

Akdeniz Language Studies Conference 2012

# The significance of data-driven descriptions of forms in explicit grammar instruction

M. Ali Bolgün\*

*Defense Language Institute Foreign Language Center, \*\* Monterey, California, 93944, U. S. A.  
Monterey Institute of International Studies, \*\* Monterey, California, 93940, U. S. A.*

---

## Abstract

Researchers have been acknowledging that explicit grammar instruction has a positive effect on learners' attainment of higher language-proficiency levels and that it is an important piece of the language teaching puzzle. One major issue in teaching lexico-grammatical elements of a language explicitly is that the descriptions of those elements are often not thorough enough to be accurately captured by a grammar rule. The descriptions in grammar books in many languages lack breadth and depth; a lot of linguistic elements may be left out, and those that are described are often not described in enough detail. This leaves students and teachers alike with unanswered questions or with rules that do not apply to all uses of a given linguistic element, both of which can be very frustrating for teachers and language learners. It is argued in this paper that descriptions of lexico-grammatical elements should be data-driven (not intuition-driven) and that corpus-linguistic analyses could help to provide actual usage-based, rather than intuition-based, descriptions and explanations to language elements. Such approach is illustrated through English and Turkish examples.

© 2012 The Authors. Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](#).  
Selection and peer-review under responsibility of ALSC 2012

*Keywords:* Linguistic corpus; corpus-based research and analysis; corpus linguistics; language teaching; language learning; explicit; implicit

---

## 1. Introduction

Of the four strands that Nation and Newton (2009) argue for (namely, 'meaning-focused input,' 'meaning-focused output,' language-focused learning,' and 'fluency development') in a language

---

\* M. Ali Bolgün. Tel.: +1-831-402-0338 Fax.: +1-831-242-6117  
E-mail address: [alibolgun@gmail.com](mailto:alibolgun@gmail.com)

\*\* The views expressed in this article are those of the author and do not necessarily reflect the official policies of the Defense Language Institute Foreign Language Center or the Monterey Institute of International Studies. No funding was received for the research reported in the article.

program, language-focused learning has to do with learning (as opposed to acquiring) of language elements, such as grammar, spelling, and pronunciation. They recommend that relatively an equal amount of time should be spent on each of these four strands. Nation and Newton are not the only ones who argue for a deliberate focus on grammar, at least as one of the components of a language program. Other researchers also acknowledge that while communicative activities and fluency are important, they are not sufficient for language acquisition, and that explicit instruction is also an important piece of the language teaching puzzle (see for example, DeKeyser, 1998; Ellis, 1998; Muranoi, 2000; Spada, Lightbown, & White, 2005; Swain, 1998; among others). In fact, even being present in the target culture does not guarantee that the individual will achieve a high level of proficiency. Angelelli and Degueldre (2002) argue that at what they call the ‘superior’ or ‘distinguished proficiency’ level, “simply spending time abroad is not necessarily sufficient for their more specialized needs. [Learners] do not need just exposure; they need answers to questions and explanations that they can rarely get by simply being immersed in a language/culture” (Angelelli & Degueldre 2002).

One major difficulty in teaching language elements (specifically grammar, and vocabulary) is that the descriptions of those elements are often not thorough. The descriptions that are available lack breadth and depth; a lot of linguistic elements (and cultural elements, for that matter) may be left out, and those that are described are often not described in enough detail (cf. Hawkins, 1984). This leaves students and teachers alike with unanswered questions or with rules that do not apply to all uses of a given linguistic element. The reason behind this is most likely because grammar books are usually written by grammarians who are native speakers of the language and the description of grammar is based on their own native-speaker intuitions. The general assumption by many native speakers is that we (as native speakers) know ‘about’ our languages when in fact our intuitions about, for example, two or more, seemingly identical, language structures may not be correct (Malmkjaer 2004; Wolfson 1989). Explanations based on our intuitions may leave many questions by students (especially at advanced levels) unanswered. Language is typically a subconscious process so while native speakers correctly choose between two or more seemingly identical structures, such as *I like to play soccer* and *I like playing soccer*, or modals, such as *should*, *ought to*, *have to*, and *must*, or words such as *uninterested* and *disinterested* and use them in proper contexts, they are not necessarily aware of the differences between them. Similarly, native speakers of English, for example, undoubtedly use correctly, and in proper contexts, the verbs *fix*, *mend*, and *repair*, and native speakers of Turkish the lexico-grammatical elements *-diği halde* and *-mesine rağmen* (both of which roughly mean ‘despite...’) and yet they may not (and most likely do not) know at the conscious level the difference in meaning, or the distribution of their use. Moreover, even in situations where their intuitions are reliable, native speakers often cannot formulate the rule(s) regarding the language issue they are asked about (Haegeman and Guéron 1999). Put differently, native speakers know the language but they may not know *about* the language. This is a major reason behind many language instructors’ being in distressing or awkward situations in which students ask the difference between, for instance, two structures to which they (i.e., teachers) do not have an answer that consistently applies to all uses of those structures, and usually respond with “*they are similar*,” “*they are interchangeable*,” “*they are based on personal preference*,” or simply “*I don’t know*.” Not finding answers to their questions is equally frustrating for language learners (cf. Byrnes, 2006).

For attainment of advanced levels of proficiency, it is especially important to know ‘about’ the language we are teaching because advanced levels of proficiency require speakers to know those subtleties of the language they are speaking. To give a few examples from Leaver and Shekhtman (2002), advanced speakers know precisely how to say what they want to say [appropriateness of expression]. They know the “rules” of the language. For example, they know the difference between *the Simple Past* and *the Present Perfect*, the difference between *mend*, *repair*, and *fix* [linguistic competence]. Their use of vocabulary is non-compensatory; if they mean “principle” they say so; they do not

compensate it with, for example, “boss” [precision of lexicon]. They understand extended discourse with knowledge (and application of) various genres, are ready to participate in conversations, know when to start a conversation, or when to be silent. They know when and how to express their emotions [discourse, emotional, and social competence]. To sum up the list, advanced learners know when and where to say exactly what to whom.

It is argued here that in the effort to help learners achieve higher levels of proficiency, precise descriptions of grammar structures, lexicon, and sociolinguistic elements need to be available to both teachers and learners. It is also argued that in order to achieve this, descriptions should be data-driven and based on analyses of language elements in context, rather than being intuition-based, and that linguistic corpora could greatly facilitate this process.

## 2. Describing Language Features

How do we come up with answers to questions we have about language? How do we uncover patterns in speakers' choices? For example, why do native speakers of English prefer *should* over *ought to* in some contexts and *ought to* over *should* in others? What are the patterns for *fix*, *mend*, and *repair*? *Help* and *assist*? *Analogous* and *similar*? Such synonymous-looking lexical or lexico-grammatical pairs are abundant in languages. For example, what is the difference between *lazım* and *gerek*, both of which roughly mean ‘need(ed)’ in Turkish? For example, *Gitmem gerek/lazım* ‘I need to go’ [lit. ‘My going is needed.’] If they absolutely mean the same thing, what, then, is their distribution? Is one used with past events more often than not? Do men or women prefer one over the other? Is it generational? Do younger people use one of them more than the other? The meaning and the patterns of use of such word pairs, expressions, and grammar structures need to be discovered and described methodically and scientifically because descriptions based on only native speaker intuitions, even if they turn out to be correct, may leave out many aspects of those language elements.

The most scientific way of uncovering patterns and coming up with explanations is to look at numerous instances of the ‘problematic’ language elements in context. To do this, however, we first need to find instances of the words, expressions, and structures that we want to know more about. The best way of doing this is to collect samples from naturally-occurring written and spoken texts, and analyze those samples in order to uncover linguistic patterns. The effort to find samples is greatly facilitated when a linguistic corpus is used simply because an overwhelmingly large percentage of words, expressions, and many grammar structures have a very low frequency of occurrence in a naturally occurring discourse (Nation, 2001). For example, think about how many times you have encountered the word *vagaries* in English. The chances are you have seen it only a few times, if at all, or perhaps never. Yet, as a learner of English, when you see this word, you may want to know more about its meaning and use. Now imagine how long it would take to manually (without the help of a linguistic corpus) find enough instances of this word's use in context. It would perhaps take days, if not weeks and months. Yet, it would take only a few seconds to do so with the help of a linguistic corpus. One must note here that learners of English are fortunate in that English is one of the lucky few languages that have been studied thoroughly. As such, learners of English are likely to find answers to such questions much more readily. However, this is not the case for the majority of the languages.

### 2.1. Linguistic corpora, concordancers, and concordance

A linguistic corpus is a collection of text samples compiled from various sources, and is basically a large collection of text often saved as a text-file. These texts are systematically selected to reflect the

language use in society. A well-balanced corpus is like a microcosm of the language it represents. Doing research using a well-constructed linguistic corpus would almost be the same as doing research involving all speakers of a language (Biber, Conrad, & Reppen, 1998). Annotated, tagged, part-of-speech tagged, and lemmatized corpora would have every single word in the corpus ‘tagged’ (labeled), thus enabling the user to establish criteria while doing searches. That way, the user can specify the part of speech (verb or noun) of, for example, the word record instead of finding all instances of any part of speech (verb and noun). With a concordancer, finding words, expressions, and structures in a well-built corpus takes only a few seconds. A concordancer is software that produces concordance. Concordance is a list of KWIC—Key Word In Context; a list of instances of a word in its immediate context. Concordance outputs make patterns more noticeable (see Figure 1). In Figure 1, the concordance shows *day by day* in context. Each line may have come from a different part of the corpus that is being utilized and is not necessarily related to the previous or the next line. Concordancers typically provide users with the capability to determine how many words before and how many words after the word (or any other linguistic element being searched) they want to see. This provides control of the linguistic context and helps with the analysis of the language element being explored. Examples (e.g., *day after day* and *day by day*) of how linguistic context helps with the analysis are given below.

Fortunately for the language professionals, nowadays, major corpora come with built-in concordancers, ready to do searches. This eliminates (for the most part) the need to obtain a separate concordancer and to learn how to use it (granted that concordancers are already simple programs to use, relatively speaking). Corpus of Contemporary American English is one of such corpora (See, for example, Fig. 1).

## 2.2. What can be searched using corpora? Some examples

Below are some examples that show how a linguistic corpus can help with the description of language features and reduce or eliminate intuition-based explanations.

The meaning and distribution of words (e.g., *fix*, *mend*, *repair*), grammar structures (e.g., *should* vs. *ought to*), phrases/idioms (e.g. *if need be*), discourse (e.g., anaphoric and cataphoric reference), registers (e.g., formality vs. informality), among others can be researched with the help of corpora (Biber et al., 1998). These are exactly the elements that are needed for a learner to reach advanced level proficiency (see the reference made to Leaver and Shekhtman, 2002, earlier). (Areas such as second language acquisition, and historical linguistics, also benefit from corpus research. However, these areas often require specialized corpora. For example, second language acquisition research would require corpora which are compiled using second-language-learner language, including grammar, spelling, and pronunciation errors, among others [see, for example, Borin and Class, 2002; Chipere, Malvern, and Richards, 2002; Nesselhauf, 2002; Tono, 2002, for specialized corpora]). Below is an example of how a linguistic corpus can be used to help us discover the meaning and distribution of two seemingly identical expressions *day after day* and *day by day*. Please note that these two expressions are from Tsui (2004) but the analyses of the expressions as presented here, including any errors they might contain, are mine. Native speakers of English use these expressions perfectly well, in proper contexts but they may not (and usually cannot) tell you why they picked one over the other in a given context. Their intuitions about the meaning and/or the distribution (when and where a given item is used) may or may not be accurate. A linguistic corpus search gives us a chance to either find out the reason behind those choices or confirm those intuitions. Using the 425-million-word Corpus of Contemporary American English (COCA) (<http://corpus.byu.edu/coca/>), let us type the string *day by day* in the Search box and choose KWIC for display (at the top left of the screen). We, then, instantly get 612 tokens of *day by day* shown in the KEYWORD IN CONTEXT DISPLAY area (see Fig 1).



When we take a look at the immediate linguistic context (the words that come before and after the expression *day by day* in each line) in the concordance window, we see that *day by day* denotes, to a large extent, neutral or positive experiences and used with words and expressions, such as *improve*, *slowly but surely*, *realistic*, *see*, *apparent*, and *courage*, etc. When we repeat the above steps for *day after day*, and look at the immediate linguistic context, we see that *day after day* denotes negative experiences and is used with words, such as *victims*, *corpses*, *inconsolable grief*, *clashing*, and *painful* among others.

he was doing , but I was n't following it	day by day	so I could pin down and say this is what
etter . But the worst was yet to come .	Day by day	Tammy seemed to grow less attentive to Andre
: from the leader of a country . He said	day by day	the role of these arms terrorists , as he called
. We visited the shop daily for a week .	Day by day	their bowl of water evaporated . It turns out tha
g any of the withdrawal symptoms and	day by day	they gain their strength and within several days
ourney the way a river seeks the sea ,	day by day	traveling into the distance , not a sudden partin
: of years finally ! finished , tracking as	day by day	we approach that apocalypse . " There was a mc
ie 1997- ' 98 El Nino , we could tell you	day by day	what was happening . " The reason ? Two new
with a small piece of wood or a metal .	Day by day	you will notice that the beautiful grass that is full
going on set . The script was " rewritten	day by day	" as shooting progressed . # Although he once h
ong the short path ( which grew longer	day by day	) that connected his cave to the road . It
Duh ! " so he went on to say , "	Day by day	across billions of ad impressions , this makes a
I loved Robert Kincaid . I dealt with it	day by day	all these years , just as he did .
e and more women trekked to the well	day by day	and as they were the majority of church-goers
the gouge in the Pentagon has widened	day by day	as the damage to the nation 's symbolic fortres
vents was known rather than recorded	day by day	as they were happening . The author had had
m , he said , " You go through your life	day by day	but when you see such a beautiful place --
the 1990s . " The rest , we eat and live	day by day	by the grace of God . We have only
ood trees that were giving up summer	day by day	crouched in the dawn with leaves already black
this miracle of mercy . May our walls	day by day	fall before his merciful onslaught . The astonish

Fig. 1. Screen shot from the Corpus of Contemporary American English (<http://corpus.byu.edu/coca/>), showing the results for day by day.

Another example, this time from Turkish, relates to the demonstratives and their referents. Turkish has three basic demonstratives, namely *bu*, *şu*, and *o*. To native speakers, and grammar books (see, for example, Banguoğlu, 2004; Gencan, 2001, and Kornfilt, 1997, among others) the fundamental difference is that *bu*, like 'this' in English, is used to refer to entities that are in very close proximity, *şu*, like 'that' in English, is used to refer to entities that are farther away, and *o* is used to refer to entities that are even farther than those referred to by *şu*. However, with *bu* and *o* both referring to what was mentioned earlier, it is not clear what the difference is (cf. Göksel & Kerslake, 2005). To many native speakers, they can be used interchangeably. A closer examination, however, indicates that demonstratives in Turkish follow a highly predictable pattern (Bolgün, 2004). *Bu* refers back to an NP used in either in the same or the previous sentence or in the second preceding sentence. Only in a very few instances, *bu* refers to NP antecedents used in the third, fourth, or the fifth preceding sentences. In some cases, as in the plural of *bu*

(*bunlar*), *bu* refers to the totality of things (to an overall idea) that were mentioned in the preceding few sentences. In any case, *bu* always refers back; never forward. For instance, in the example below, *bu* refers to the underlined NPs in the preceding sentence:

(1)  
Hayır, sadece bir yaşam sevgi-si-yle, bir yaşam zevk-i sorun-u-dur. **Bu**, çok  
no only one life love-CM-WITH one life pleasure-CM problem-POSS-COP this much  
  
önemli mi-dir?  
important Q-COP

<<http://www.milliyet.com/2003/06/02/yazar/altan.html>>

‘No, (it is) only a matter of love of life and a joy of life. Is **this** very important?’

Of the total 102 uses of *bu* that were found and analyzed, 73 (or 71.5%) are used anaphorically; to refer back to an antecedent (or antecedents). 29 (or 28.5%) of the total 102 are used in situations where the referent is contextually present, or as part of an adverbial phrase, as in *bu ara-da* [this gap/time-LOC] ‘meanwhile’. Further, of the total 73 anaphoric uses of *bu*, 61 refer to antecedents present in the same or the previous sentence, 5 of them in the second preceding sentences, 3 of them in the third preceding sentences, 2 of them in the fifth preceding sentences and, 2 of them refer only to the totality of what was mentioned in the previous five or more sentences.

Parallel to *bu*, *o* (traditionally translated as ‘he,’ ‘she,’ ‘it,’ or ‘that’) also has the anaphoric and situational uses. However, *o* differs from *bu* both in terms of number of uses and in terms of the nature of uses. There were 32 instances of *o* as opposed to 102 instances of *bu*. The numbers indicate that while there are a considerable number of uses of *o*, *bu* is favored in more instances than *o*. The examples were collected from an online version of a Turkish newspaper *Milliyet* and it could be that journalists want to sound more accurate and talk or write about matters that are more tangible, clearer, closer in time, more ‘still relevant’ (rather than ‘not-anymore relevant’) situations unless they have to. *O* often provides the opposite: the antecedents of *o* are often abstract, hypothetical, or farther away in time; the boundaries are less clear. The following example illustrates this:

(2)  
Mezun ol-an genç-ler Türkiye’de iş bulabilme konu-su-nda hayli umutsuz. **O**  
graduate be-NOM youth-PL Turkey-LOC job finding topic-CM-LOC quite hopeless that  
  
neden-le ‘biz-i kaybed-iyor-sunuz’ de-yip yurtdışı-nda gelecek arı-yor-lar.  
reason-WITH we-ACC lose-PROG-2PL say-ADV abroad-LOC future search-PROG-PL

<http://www.milliyet.com/2003/05/19/siyaset/asiy.html>

‘The young (people) who are university graduates are fairly hopeless in regards to finding jobs. **That** is why they say ‘you are losing us’ and look for a future abroad.’

In the above example, it is grammatically possible to use *bu* (as in *bu nedenle* ‘that is why’) instead of *o*. However, that choice (*bu*) would have given the impression that the idea presented in the first sentence

of the example is that of the writer. By choosing *o*, the writer chooses to somewhat distance himself and to express the same idea from the ‘young people’s perspective, bringing objectivity to his argument.

Unlike *bu* and *o*, the demonstrative *şu* is used to refer forward (cataphorically). With a total of only 9 instances, the number of uses of *şu* is much less than both *bu* and *o*. In other words, out of a total of 143 demonstratives, only nine (or 6.2%) of them are *şu* ‘that’. Of the nine instances of *şu*, four (or 44.4%) refer cataphorically while five (or 55.5%) of them are situational uses. The following illustrates the typical cataphoric use of *şu*:

(3)

Sanayi-de ise nitelikli insan güc-ü el-de ed-e-me-dik-ler-i için  
industry-LOC though qualified human power-CM hand-LOC make-ABIL-NEG-NOM-PL-POSS for

dialog kopukluğ-u-ndan söz ed-il-ir... Bekle-nil-en **şu**-dur:  
dialog disconnection-CM-ABL mention do-PASS-AOR expect-PASS-NOM that-COP

Sanayi-nin iste-yeceğ-i tarz-da insan yetiştir-mek.  
industry-GEN want-FUT.NOM-POSS style-LOC human train-INF

<<http://www.milliyet.com/2003/05/19/siyaset/asiy.html>>

‘In the industry, though, since they are not able to attain qualified work force, they talk of a lack of dialogue... What is expected is **this**: Training the kind of people that the industry would want.’

The translation of *şu* in the above example is given as ‘this,’ which is traditionally assigned to *bu*. However, this is only a matter of translation. What is important here is that *şu* points forward to what is going to come next.

Another example of how corpus analysis helps to discover patterns relates to the words for ‘privacy’ and ‘private’ in Turkish. In their corpus-based study on the emergence of words related to privacy in Finnish and Turkish, Kuha & Bolgün (2009) show that while there are a number of Turkish words that can be translated with ‘private’ or ‘privacy’ (see, for example, Akdikmen, Uzbay, & Özgüven, 2006), there is a clear pattern with regards to their distribution. Of the two most frequently used words, *mahrem* ‘private’ is used in highly intimate situations and for sexually-charged expressions, such as *mahrem yerler* ‘private parts,’ referring to sexual organs, whereas *özel* ‘private’ is used (often with *yaşam* ‘life’) in other contexts.

The majority of the instances of *özel* found in the corpus center around the meanings of ‘special’, ‘peculiar’, ‘specific’, ‘unique’, and ‘privileged’. Consider the following example from METU Turkish Corpus (Say, Zeyrek, Oflazer, & Özge, 2002). (Hereafter, unless otherwise indicated, the examples provided were obtained from METU Turkish Corpus (Say et al., 2002).

(4)

Eğer sıradan bir vatandaş-sa-nız, demokrasi-yle ilgi-niz ol-ma-z. Eğer **özel** bir  
If ordinary a citizen-IF-2PL democracy-WITH concern-2PL be-NEG-AOR If special a

vatandaş-sa-nız, o zaman en tehlikeli iş-ler-e de giriş-se-niz  
citizen-IF-2PL that time most dangerous job-PL-DAT also venture-IF-2PL

demokrasi sizin için çalışır.  
 democracy your for work-AOR

‘If you are an ordinary citizen, you would have nothing to do with democracy. If you are a **special** citizen, then democracy works for you even if you undertake dangerous businesses.’

It is clear that by *özel bir vatandaş*, what is meant is ‘a privileged citizen’ and not ‘a private citizen’. This is made clear by the adjective *sıradan* ‘ordinary’ that is used in the first sentence. The second sentence contrasts such a citizen (an ordinary one) with a non-ordinary one.

In the example below, *özel* indicates that noun phrase it modifies is a ‘non-government’ entity.

(5)

*Hangisi zor, kamu görev-i mi, özel sektör mü?*  
 Which one difficult government job-CM Q private sector Q

‘Which one is more difficult: government job, (or) **private** sector?’

In 10.7% of the instances, *özel* simply indicates that the noun phrase it modifies belongs to the person; it means *personal*. There is nothing necessarily private about it. For example:

(6)

*Bu da Cumhurbaşkanı-nın özel merak-ı-nı gider-mek için yap-tığ-ı*  
 this also president-GEN personal curiosity-POSS-ACC quench-INF for do-NML-POSS

*gezinti-ler-in inceleme olarak nitelendir-il-me-si-dir.*  
 tour-PL-GEN survey as characterize-PASS-NML-POSS-COP

‘And this is characterizing the trips that the president takes to satisfy his **personal** curiosity as being (official) surveys.’

In 1.6% of the instances, *özel* is simply a last name, a company name, or part of a book title.

Of the remaining 48 instances (or 16.0%) of *özel* (to mean ‘private’), 27 (or 56.2%) collocate with *yaşam*, *yaşantı*, or *hayat* all of which mean ‘life’.

Meanwhile, all instances of *mahrem* and *mahremiyet* mean ‘private’ or ‘privacy’ in the most intimate way. It seems that although *özel* has taken over a number of borrowings (such as *hususî* ‘special’) to a large extent, there are some areas where *özel* does not quite express the intended meaning. Even when *özel* is used with *yaşam* ‘life’, it often means ‘non-job related life’ and not ‘private life’ in its strictest sense to mean ‘no access’ to that characterized as *özel*. For example:

(7)

*Fakat Cumhurbaşkanı da benim gibi bir insan. Konuşmayı sev-en bir insan.*  
 however president also me like a person talk-ACC like-NML a person



*Duy-duğ-u-m-a göre özel yaşam-ı-nda çok hoşsohbet-miş.*  
 hear-NML-POSS-1SG-DAT according private life-POSS-LOC very sociable-PAST

‘However, the president, too, is a human being too; just like me. A human being that likes to talk. As far as I have heard, he is very sociable in his **private/personal** life.’

Similarly, the following contrasts ‘private/personal life’ with ‘job life’:

(8)  
 Şekerim, insan özel yaşam-ı-yla iş-i-ni ayır-malı. Ne ev-e  
 honey person private life-POSS-WITH job-POSS-ACC separate-OBLIG neither home-DAT

taşı-malı-sın, ne de ev-in-i iş-in-e.  
 carry-OBLIG-2SG nor also home-POSS-ACC job-POSS-DAT

‘Honey, one should separate one’s **personal/private** life from one’s job. You should not bring your job home; neither should you bring your home to your job.’

*Mahrem*, on the other hand, is stricter in that, access or penetration to that characterized by *mahrem* constitutes a bigger violation of the norm. This is not surprising since *mahrem* shares the same root with the words that mean ‘forbidden,’ ‘prohibited,’ ‘unlawful,’ ‘sacred,’ and ‘sin,’ among others (Ba‘albaki, 1996; Cowan, 1994). Compare the two examples above with the following two examples:

(9)  
 Vücut-u-nu şehvet düşkünlüğ-ü-yle öylesine kötü-ye kullan-mış-tı ki,  
 body-POSS-ACC lust addiction-CM-WITH such bad-DAT use-PPTCL-PAST CONJ

*mahrem* yer-ler-i başka kadın-lar-ın-ki gibi doğa-nın belirle-diğ-i  
 private part-PL-POSS other woman-PL-GEN-REL as nature-GEN determine-NML-POSS

yer-de değil-di ve sanki yüz-ü-ne vur-muş-tu  
 place-LOC not-PAST and as if face-POSS-DAT reflect-PPTCL-PAST

‘She had used her body with lust so badly that her **private** areas were not where the nature intended for them to be, like in other women, and it is as if (her lust) was reflected in her face.’

In the example above, what is meant by private areas are clearly sexual organs. In such a context, *mahrem* is picked over *özel*. Similarly, in the example below, the context is lovemaking and the adjective used in this context is *mahrem*.

(10)  
 Para-sı-nı ver-ip sokak-lar-dan sahip-siz beden-ler topla-mak, onlar-la bu boş  
 money-POSS-ACC give street-PL-ABL owner-less body-PL collect-INF they-WITH this empty

ev-in mağara kovuk-lar-ı-na benze-yen sessiz oda-lar-ı-nda *mahrem*  
 house-GEN cave hole-PL-CM-DAT resemble-NML quiet room-PL-POSS-LOC private

oyun-lar oyna-mak bir mucize gibi gel-iyor-du ban-a; orospu-lar-la yaşa-dığ-ım  
game-PL play-INF a miracle as come-PROG-PAST I-DAT prostitute-PL-WITH live-NML-1SG

her parçala-n-mış sevişme-den sonra büyük bir huzur ve ferahlık  
every break-PASS-PTCL lovemaking-ABL after big a peace and contentment

duy-uyor-du-m.  
feel-PROG-PAST-1SG

‘Paying for and collecting ownerless bodies from the streets and playing with them **private** games in this house’s rooms that looked like hollows of caves seemed like a miracle to me; I was feeling a sense of peace and contentment after every shattered lovemaking that I had with prostitutes.’

The high number of instances of *özel* being used with ‘life’ could be due to the lack of a Turkish noun that means *privacy*. *Mahremiyet*, the Arabic borrowing meaning ‘privacy’, does not always satisfy the current need since it refers to a specialized form of privacy. It appears that the recent *özel yaşam* and the older *mahremiyet* are both needed to compensate for *privacy*. For example, *özel yaşam* ‘private life’ is too general in the context below and cannot substitute for *mahremiyet* ‘privacy’. Consider:

(11)  
Verimliliğ-i önemse-yen kimi şirket-ler-in ofisleri adeta kusursuz-du. Ama  
productivity-ACC value-PTCL some company-PL-GEN office-PL-POSS almost flawless-PAST but

hiçbir-i-nde çalışan-lar-ın mahremiyet-i-ne önem ver-en bir tasarım  
none-POSS-LOC employee-PL-GEN privacy-POSS-DAT importance give-NML a design

yoktu.  
nonexistent-PAST

‘The offices of some companies that value productivity were almost flawless. However, in none of them was there a design that values the employees’ **privacy**.’

In the above example, the issue under discussion is not employees’ private lives; for example, what they do in their own time. Rather, it is the way the workplace is set up and how, perhaps, the employees are exposed to other people’s gaze in that setup.

In another corpus-based study, Bolgün (2005) shows that definiteness, specificity, and referentiality cannot explain the meaning and function of the Turkish accusative case, a topic of interest in linguistic literature on Turkish for some 340 years (Seaman 1670), if not more. The direct object (DO) in Turkish has four distinct types. These are illustrated in boldface in the following four examples (taken from Taylan and Zimmer 1994).<sup>†</sup>

(12)  
(a) Ali her gün **gazete-yi** oku-yor.  
Ali every day **newspaper-ACC** read-PROG

<sup>†</sup> Boldfacing is added; the gloss of the first example is slightly modified from the original, and glosses have been added to examples (2), (3), and (4).

‘Ali reads the newspaper everyday.’

- (b) Ali her gün **bir gazete-yi** oku-yor.  
Ali every day **one newspaper-ACC** read-PROG

‘Ali reads a newspaper everyday.’

- (c) Ali her gün **bir gazete** oku-yor.  
Ali every day **one newspaper** read-PROG

‘Ali reads a newspaper everyday.’

- (d) Ali her gün **gazete** oku-yor.  
Ali every day **newspaper** read-PROG

‘Ali reads a newspaper/newspapers everyday.’

The boldfaced nouns in the above examples share a common feature: they all occupy the unmarked DO position, immediately before the verb. What is different about these DOs is that (a) has the accusative (ACC) marker *-(y)I*, (b) has the ACC marker and is preceded by *bir* ‘one,’ (c) is also preceded by *bir* ‘one’ but does not have the ACC marker, and (d) is in its so-called *bare* form; it neither has the ACC marker nor is it preceded by *bir* ‘one.’ (Please note that Taylan and Zimmer (1994) use the term ‘indefinite article’ to refer to *bir* ‘one.’ However, there is no consensus on this. For example, while Kornfilt (1997), Lewis (2000), Swift (1963), Taylan and Zimmer (1994), and Tura (1973) treat it as such in certain uses, others do not).

Given these different ways of expressing the (seemingly) same idea, the question arises as to what the difference is. Because Turkish does not have any morphological determiners or a definite article, such as *the* in English (e.g., Erguvanlı, 1984; Kornfilt, 1997; Underhill, 1976), the accusative case, one of the six cases in Turkish, has traditionally been characterized (generally speaking) either as corresponding to the definite article in English (e.g., Ergin, 1962; Erguvanlı, 1984; Erguvanlı-Taylan, 1987; Lewis, 2000; Sebüktekin, 1971), as indicating referentiality (e.g., Dede 1986), or as indicating specificity (e.g., Aissen, 2003; Enç, 1991; Erguvanlı, 1984; Kornfilt, 1997; Swift, 1963). While these characterizations are correct to a certain extent, Bolgün (2005) shows, providing examples found in METU Turkish Corpus (Say et al., 2002), and examples collected from the online version of a Turkish newspaper, that traditional notions of ‘definiteness,’ ‘referentiality’ and ‘specificity,’ which are very commonly thought of as being indicated by the accusative case marking, cannot fully account for its meaning and function.

The noun in DO position with the ACC marker but no preceding *bir* ‘one,’ as in (a) above, is generally considered to be definite, in the sense that the hearer knows or can identify the *gazete* ‘newspaper’ being mentioned. However, consider the following example, taken from the aforementioned corpus:

- (13)  
Artık **üniversite-yi** bitir-mek ve aynı kariyer-de ilerle-mek çalışma yaşam-ı  
anymore **university-ACC** finish-INF and same career-LOC progress-INF work life-CM  
  
açı-sı-ndan                      garantili    bir yol değil.

viewpoint-CM-ABL guaranteed one way not

<<http://www.milliyet.com/2003/05/19/siyaset/asiy.html>>

‘Graduating from a/the university and progressing in the same career is not a reliable way for work life anymore.’

In the example above, *üniversite* ‘university’ has the ACC. However, unlike what is claimed in some grammar books, it is not definite. What is referred to with the use of this noun that has the ACC is not a particular university that both the speaker and listener can identify. It is used generically.

DOs, with or without the ACC marker, but preceded by *bir* ‘one’ (see examples a and b above) are explained by appealing to the notion of ‘specificity,’ in the sense that a specific noun will have ‘a certain’ reading. Therefore, DOs bearing ACC (as in (b), for example) are considered specific, whereas DOs not bearing ACC (as in (c), for example) are considered nonspecific (Enç 1991). However, there are numerous examples that challenge this account. Consider the one below.

(14)

Kilise-den çok **bir ev-i** andır-an eski yapı-nın bahçe-si-nde biz  
church-ABL rather **one house-ACC** resemble-NML old structure-GEN garden-POSS-LOC we

biz-e-ydi-k işte.  
we-DAT-PAST-1PL here

‘There we were, by ourselves in the garden of the old structure that resembled a house rather than a church.’

In the example above, the noun *ev* ‘house’ has both the ACC and is preceded by *bir* ‘one’ and should be considered specific by some accounts; the noun *ev* ‘house’ should have ‘a certain’ reading. However, ACC-bearing *bir ev* ‘one house’ does not refer to any particular house that the speaker (or the hearer) knows, and cannot be said to be specific in that sense. The speaker is simply stating that the structure, the garden of which they happen to be in, resembles a house (any house, in fact) rather than a church (any church). He is comparing the structure with two entities (‘a house’ and ‘a church’), and argues that it resembles a house more than a church.

### 3. Conclusions

Achieving higher levels of language proficiency (especially, accuracy) necessitate, at least partially, knowledge of subtle distinctions between the seemingly identical structures and vocabulary in the target language. Only a handful of languages, other than English, can claim that they have a significant number of resources that provide data-driven descriptions of those language structures and vocabulary. Most languages arguably lack such descriptions (cf. Fotos, 2002). Instead, descriptions of language elements are largely intuition-based and fail to capture all their possible uses and various nuances of their meaning.

Yet, languages are abundant with features that are seemingly (and deceptively) synonymous. For example, what is the difference between *gerilim* and *gerginlik*, both of which mean ‘tension’ in Turkish? What about the difference between these three structures: *-mektense*, *-mek yerine*, and *-eçeği+PE+(n)e* all of which roughly mean *rather than* in Turkish? If language programs are to help students attain high levels of proficiency, then language elements like these need to be better described and explained without

relying on (only) intuition but rather on data and facts obtained from naturally-occurring discourse. Corpus-based research makes this possible by helping researchers analyze naturally-occurring language output efficiently. Language teaching- and reference-materials developed using the above approach with the help of linguistic corpora and concordancers virtually eliminate guess-work and explanations that are based on unreliable native-speaker intuitions.

Corpus-analytic approach to language elements and materials prepared as a result of such an approach does not and should not necessarily require any (substantial) change in the language teaching methods or techniques employed in class. Rather, accurate descriptions enhance the quality of language instruction and language learning by providing both teachers and learners with accurate answers to their questions. This is in fact what is missing from the language programs and is in itself very valuable.

That said, there have been attempts to introduce novice practices to classroom teaching based on corpus-linguistic research. For example, Johns (1994) developed the Data-Driven Learning (DDL) method, in which the learner essentially assumes the role of a researcher, accessing language elements in a language corpus via a concordancer, looking for patterns and meaning(s) of those elements. This method can lead to student autonomy, and should be encouraged to a certain extent and with advanced speakers. However, DDL is not sustainable in many language-learning situations where students are busy with other courses and obligations; they cannot be expected to find patterns and make generalizations regarding every language issue they encounter. In fact, even teachers may find it hard to allocate enough time for a corpus analysis, or find it very difficult to get into the corpus-linguistic analysis mentality (cf. Mauranen, 2002). Corpus-based analyses lead to data-driven and accurate descriptions but the process of analyzing language elements can be very time-consuming and is not suitable for extensive use in classrooms. A better approach would be for language professionals, (e.g., materials developers, including reference-books writers) and to some extent the teachers to do the bulk of the research and analysis of a given language issue and perhaps involve students afterwards through, for instance, cloze tests based on sentences obtained from the linguistic corpus. For example, instead of claiming that there is no difference, the difference between *day after day* and *day by day* can be explained using some of the sentences collected from the corpus. After the explanation and sample sentences, a cloze-test can be prepared by simply deleting the expressions (*day after day* and *day by day*) being taught from those sentences, asking the students to decide whether the missing words should be *day after day* or *day by day* in those contexts, and to state why they made those choices thereby raising their consciousness regarding those expressions.

Similarly, reference-book explanation of the Turkish demonstratives could include statistical information regarding their frequency of use and distribution, and naturally occurring examples to illustrate those uses could be incorporated.

The words for privacy can be shown using examples pulled out of the corpus, and crucial point that separates the multiple words that translate into the English 'private' or 'privacy' can be stated. This would answer students' questions in a much better way than dictionaries would because in all the available dictionaries, a number of different words in Turkish are translated with 'private' or 'privacy.'

The use of accusative does not need to be wrongly equated with concepts such as definiteness, specificity, or referentiality. When students find examples that counter teachers' explanations that involve these (or similar) concepts, it is very frustrating for them, especially when they hear something like, "In this context, we just say it this way." Instead, students could be told that while the accusative in Turkish and the definite article in English often overlap, they are not identical in meaning and function. In Turkish, the function of the accusative is individuation, the presentation of the entity denoted by the direct object noun as complete, and separate from all others that may be around it.

Alternatively, (with advanced-level students) for some of the examples above (except the ones that



may require expertise in linguistics or the language), the teacher can find the instances of the ‘problematic’ language elements (such as *day after day* or *day by day*), and then ask the advanced-level students to ‘discover’ the differences between them (cf. Ellis, 2002), thereby doing some of the time-consuming work for them.

One final word is that, if a linguistic corpus does not already exist for a given language, because the initial investment (in time, energy, and manpower) is significant, building one should be supported at least at the institutional level. Institutions with similar language programs can cooperate and speed up the process of building linguistic corpora to be used by all that are teaching or learning that language. The building of a corpus should include a built-in concordancer to make it user-friendly and to increase the chances of its use by language professionals.

## References

- Aissen, J. (2003). Differential Object Marking: Iconicity vs. Economy. *Natural Language & Linguistic Theory*, 21:3, 435-483.
- Akdikmen, R., Uzbay, E., & Özgüven, N. (2006). *Langenscheidt New Standard Turkish Dictionary*. Berlin and Munich: Langenscheidt KG.
- Angelelli, C. & Degueldre, C. (2002). Bridging the gap between language for general purposes and language for work: An intensive Superior-level language/skill course for teachers, translators, and interpreters. In, B. L. & B. Shekhtman (Eds.), *Developing professional-level language proficiency*. Cambridge University Press.
- Ba‘albaki, M. (1996). *Al-Mawrid: A Modern English-Arabic Dictionary*. Beirut, Lebanon: Dar El-Ilm Lil-Malayan.
- Banguoğlu, T. (2004). *Türkçenin Grameri* [Grammar of Turkish] (7<sup>th</sup> printing). Ankara: Türk Dil Tarih Kurumu. 1<sup>st</sup> printing: 1959.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Bolgün, M. A. (2004, November). *A poster about ‘this’ and ‘that’ in Turkish and English: some differences in how demonstratives operate in Turkish and English – a possible source of mistakes in Turkish EFL learners’ use of English demonstratives*. Poster session presented at the meeting of Indiana Teachers of English to Speakers of Other Languages [INTESOL] Conference. Carmel, Indiana.
- Bolgün, M. A. (2005). *Accusative marking in Turkish* (Doctoral dissertation). Retrieved from ProQuest. (AAT 3209425)
- Borin, L. & Prütz, K. (2002). New wine in old skins? A corpus investigation of L1 syntactic transfer in learner language. In G. Aston, S. Bernardini, & D. Stewart (Eds.). *Corpora and Language Learners*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Byrnes, H. (2006). What kind of resource is language and why does it matter for advanced language learning? An Introduction. In H. Byrnes (Ed.) *Advanced Language Learning: The Contribution of Halliday and Vygotsky*. London/New York: Continuum.

Chipere, N., Malvern, D., & B. Richards. (2002). Using a corpus of children's writing to test a solution to the sample size problem affecting type-token ratios. In G. Aston, S. Bernardini, & D. Stewart (Eds.). *Corpora and Language Learners*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Cowan, J. M. (Ed.). (1994). *Arabic-English Dictionary: The Hans Wehr Dictionary of Modern Written Arabic* (4<sup>th</sup> ed). Ithaca, NY: Spoken Language Services, Inc.

Dede, M. (1986). Definiteness and referentiality in Turkish verbal sentences. In D. I. Slobin & K. Zimmer (Eds.) *Studies in Turkish Linguistics*. Amsterdam / Philadelphia: John Benjamins Publishing Company.

DeKeyser, R. M. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42-63). Cambridge University Press.

Ellis, R. (1998). Teaching and research: Options in grammar teaching. *TESOL Quarterly*, 32, 39-59.

Ellis, R. (2002). Methodological options in grammar teaching materials. In E. Hinkel & S. Fotos, *New Perspectives on Grammar Teaching in Second Language Classrooms*. New Jersey & London: Lawrence Earlbaum Associates, Inc.

Enç, M. (1991). The semantics of specificity. *Linguistic Inquiry*, 22, 1-25.

Ergin, M. (1962). *Türk Dil Bilgisi*. İstanbul Üniversitesi Edebiyat Fakültesi Yayınları.

Erguvanlı, E. E. (1984). *The Function of Word Order in Turkish Grammar*. University of California Press.

Erguvanlı-Taylan, E. (1987). The role of semantic features in Turkish word order. *Folia Linguistica*, 21, 215-227.

Fotos, S. (2002). Structure-Based Interactive Tasks for the EFL Grammar Learner. In E. Hinkel & S. Fotos, *New Perspectives on Grammar Teaching in Second Language Classrooms*. New Jersey & London: Lawrence Earlbaum Associates, Inc.

Gencan, T. N. (2001). *Dilbilgisi* [Grammar]. Istanbul: Ayraç Yayınevi.

Göksel, A. & Kerslake, C. (2005). *Turkish: A comprehensive grammar*. Routledge.

Haegeman, L. M. V. & Guéron, J. (1999). *English Grammar: A Generative Perspective*. Blackwell Publishing.

Hawkins, E. (1984). *Awareness of language: An introduction*. Cambridge, UK: Cambridge University Press.

Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of Data-Driven Learning. In T. Odlin (Ed.), *Perspectives on Pedagogical Grammar* (293-313). Cambridge: Cambridge University Press.

Kornfilt, J. (1997). *Turkish*. London; New York: Routledge.

Kuha, M. & Bolgün, M. A. (2009). The development of Turkish and Finnish words related to privacy. In H. Pishwa, *Language and social cognition: Expression of the social mind* (127-162). Berlin: Mouton de Gruyter.

Leaver, B. L. & Shekhtman, B. (2002). Principles and practices in teaching Superior-level language skills: not just more of the same. In B. L. Leaver & B. Shekhtman, (Eds.) *Developing professional-level language proficiency* (pp.3-33). Cambridge University Press.

Lewis, G. L. (2000). *Turkish Grammar*. Oxford: Oxford University Press.

Malmkjaer, K. (2004). *The Linguistics Encyclopedia* (2<sup>nd</sup> ed.). Routledge.

Mauranen, A. (2002). Speech corpora in the classroom. In G. Aston, S. Bernardini, & D. Stewart (Eds.). *Corpora and Language Learners*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Muranoi, H. (2000). Focus on form through interaction enhancement: Integrating formal instruction into a communicative task in EFL classrooms. *Language Learning*, 50, 617-673.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nation, I. S. P. & Newton, J. (2009). *Teaching ESL/EFL listening and speaking*. New York, NY: Routledge.

Nesselhauf, N. (2002). How learner corpus analysis can contribute to language teaching: A study of support verb constructions. In G. Aston, S. Bernardini, & D. Stewart (Eds.). *Corpora and Language Learners*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Say, B., Zeyrek, D., Oflazer, K. & Özge, U. (2002). Development of a Corpus and a Treebank for Present-day Written Turkish.” In K. İmer & G. Doğan (Eds.), *Current Research in Turkish Linguistics* (pp.183-192). Proceedings of 11th International Conference on Turkish Linguistics. Eastern Mediterranean University, Northern Cyprus.

Seaman, W. [a.k.a. Gulielmo Seaman]. (1670). *Grammatica linguæ Turcicæ in quinque partes distributa*. Oxoniae [Oxford]. Excudebat Hen: Hall ... Prostant apud Edvardum Millington...

Sebüktekin, H. I. (1971). *Turkish-English Contrastive Analysis: Turkish Morphology and Corresponding English Structures*. The Hague: Mouton.

- Spada, N., Lightbown, P.M., & White, J. (2005). The importance of meaning in explicit form-focused instruction. In A. Housen & M. Pierrard (Eds.), *Current issues in instructed second language learning*. Brussels: Mouton De Gruyter.
- Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64-81). Cambridge University Press.
- Swift, L. B. (1963). *A Reference Grammar of Modern Turkish*. Indiana University, Bloomington; The Hague, The Netherlands: Mouton & Co.
- Taylan, E. E. & Zimmer, K. (1994). Case Marking in Turkish Indefinite Object Constructions. In S.Gahl, A. Dolbey, & C. Johnson (Eds.) *Proceedings of the twentieth annual meeting of the Berkeley Linguistics Society* (February 18-21, 1994.). *General Session*. Berkeley Linguistics Society.
- Tono, Y. (2002). Multiple comparisons of IL, L1 and TL corpora: The case of L2 acquisition of verb subcategorization patterns by Japanese learners of English. In G. Aston, S. Bernardini, & D. Stewart (Eds.). *Corpora and Language Learners*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Tsui, A. B. M. (2004). What teachers have always wanted to know – and how corpora can help. In J. M. Sinclair (Ed.), *How to use corpora in language teaching*. John Benjamins.
- Tura, S. (1973). *A study on the articles in English and their counterparts in Turkish*. Dissertation Abstracts International, 35 (01), 436. (UMI No. 7415877)
- Underhill, R. (1976). *Turkish Grammar*. Cambridge, Mass.: The MIT Press.
- Wolfson, N. (1989). *Perspectives: Sociolinguistics and TESOL*. Cambridge, MA: Newbury House.